

IN THE CLAIMS

1. (Currently Amended) A method of computing response time of a web server, comprising the steps of:

placing a plurality of correlation tags in data at networking and application layers during networking and application processing at the web server, wherein ~~said~~ the tags allow for later identification of ~~said~~ the data;

collecting ~~said~~ the networking and application data from ~~said~~ the networking and application layers of the web server at a correlation engine in the web server;

identifying, using ~~said~~ the correlation tags, a plurality of data corresponding to a single web event at the correlation engine in the web server;

combining ~~said~~ the data corresponding to the single web event from ~~said~~ the networking and application layers into a metric at the correlation engine in the web server; and

calculating, using ~~said~~ the metric, client perceived response time of the single web event at the correlation engine in the web server.

2. (Currently Amended) The method of claim 1, wherein ~~said~~ the event is selected from one of a web page download and a web session.

3. (Currently Amended) The method of claim 1, wherein ~~said~~ calculating step ~~is~~ performed using analytical models of response time.

4. (Currently Amended) A method of estimating ~~of~~ a perceived response time of at least one web server computing device to one or more client computing devices connected to the at least one web server computing device via a network, the method comprising the steps of:

generating and placing a session identifier (ID) as a correlation tag in each of a plurality of requests sent by a client to a web server device during application processing at the web server device, wherein ~~said~~ the correlation tags identify ~~said~~ the requests;

generating and placing a connection identifier (ID) as a correlation tag in each communication packet sent between the client and the web server device during network processing at the web server device;

identifying, using ~~said~~ the session identifiers and the connection identifiers, requests and communication packets that correspond to a single event at a correlation engine in the web server device;

combining ~~said~~ the identified requests and ~~said~~ the communication packets into a metric at the correlation engine in the web server device; and

estimating, using ~~said~~ the metric, client perceived response time of the single event of ~~said~~ the ~~at least one web server computing device~~ to a request by ~~said~~ the ~~one or more client computing devices~~ connected to the web server device via a network, at the correlation engine in the web server device.

5. (Original) The method of claim 4, wherein the network is the Internet.

6. (Currently Amended) The method of claim 4, wherein ~~said step of~~ generating and placing the session ID ~~further comprises a step of~~ establishing a web session between the client and the web server device.

7. (Currently Amended) The method of claim 4, further comprising ~~a step of~~ logging each web session between the client and the web server device.

8. (Currently Amended) The method of claim 4, wherein ~~said step of~~ generating and placing ~~said~~ the connection ID ~~further comprises a step of~~ establishing a network connection between the client and the web server device.

9. (Currently Amended) The method of claim 4, further comprising ~~a step of~~ logging ~~said~~ each communication packet sent between the client and the web server.

10. (Currently Amended) The method of claim 4, further comprising ~~a step of grouping~~ all of ~~said~~ the plurality of requests and ~~said~~ the communication packets corresponding to ~~a~~ the single event.

11. (Currently Amended) The method of claim 4, wherein ~~said~~ estimating ~~step~~ further comprises the steps of:

- a) retrieving a page composition vector and TCP/IP round trip time (RTT), packet loss rate, and average connection time T_c ;
- b) calculating time $T_1 = T_c + C_1(b_1)$ and time $T_2 = C_1(o) + T_c + C_2(b_2)$ and setting a loop counter;
- c) averaging T_1 and T_2 by $(T_1 + T_2)/2$ and terminating processing if the loop counter is less than or equal to the value n ;
otherwise, if T_1 is smaller than T_2 , T_1 is set to $T_1 + (RTT/2) + C_1(BI)$, and
if T_1 is not smaller than T_2 , T_2 is set to $T_2 + (RTT/2) + C_2(BI)$; and

- d) incrementing the loop counter and repeating step c,
wherein the page composition vector is composed of $\{b_1, b_2, \dots, b_n\}$ and o ,
 b_i , where i is an number 1, 2, ..., being the size of the i -th component of the web page,
 n being the number of components,
 o being the offset at which first component is embedded in a container page,

$C_1(y)$ being the time it takes to download y bytes on a first TCP/IP connection between the client and the web server device, and

$C_2(y)$ being the time it takes to download y bytes on a second TCP/IP connection between the client and web server device.

12. (Currently Amended) An apparatus for estimating a perceived client response time of at least one web server computing device to one or more client computing devices connected to the at least one web server computing device via a network, the ~~method~~ apparatus comprising:

a means for generating and placing a session identifier (ID) as a correlation tag in each of a plurality of requests sent by the one or more client computing devices to the at least one web

server computing device during application processing at the at least one web server computing device, wherein-said the correlation tags identify-said the requests;

a means for generating and placing a connection identifier (ID) as a correlation tag in each communication packet sent between the one or more client computing devices and the at least one web server computing device during network processing at the at least one web server computing device;

means for identifying, using-said the session identifiers and connection identifiers, requests and communication packets that correspond to a single event in the at least one web server computing device;

a means for combining-said the identified requests and-said the communication packets into a metric in the at least one web server computing device; and

a means for estimating, using-said the metric, the perceived client response time of the single event of the at least one web server computing device to a request by the one or more client computing devices connected to the at least one web server computing device via the network, in the at least one web server computing device.

13. (Currently Amended) A computer program device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for an apparatus for estimating a perceived client response time of at least one web server computing device to one or more client computing devices connected to the at least one web server computing device via a network, the method comprising program executing the following steps:

~~a means for~~ generating and placing a session identifier (ID) as a correlation tag in each of a plurality of requests sent by the one or more client computing devices to the at least one web server computing device during application processing at the at least one web server computing device, wherein-said the correlation tags identify-said the requests;

~~a means for~~ generating and placing a connection identifier (ID) as a correlation tag in each communication packet sent between the one or more client computing devices and the at least one web server computing device during network processing at the at least one web server computing device;

~~means for identifying, using said~~ the session identifiers and connection identifiers, requests and communication packets that correspond to a single event in the at least one web server computing device;

~~a means for combining said~~ the identified requests and ~~said~~ the communication packets into a metric in the at least one web server computing device; and

~~a means for estimating, using said~~ the metric, the perceived client response time of the single event of the at least one web server computing device to a request by the one or more client computing devices connected to the at least one web server computing device via the network, in the at least one web server computing device.